



Audio Engineering Society

Convention Paper 10261

Presented at the 147th Convention
2019 October 16–19, New York, USA

This paper was peer-reviewed as a complete manuscript for presentation at this Convention. This paper is available in the AES E-Library, <http://www.aes.org/e-lib>. All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

An attempt to elicit horizontal and vertical auditory precedence percepts without pinnae cues.

Wesley A. Bulla¹ and Paul Mayo²

¹ Belmont University, Nashville TN

² University of Maryland, College Park MD

Correspondence should be addressed to Author (wesbulla@comcast.net and mayopg@cox.net)

ABSTRACT

This investigation was a continuation of AES-143 paper #9832 and AES-145 paper #10066 where reliable auditory precedence in the elevated, ear-level, and lowered horizontal planes was examined. This experiment altered and eliminated the spectral influences that govern the detection of elevation and presented two different horizontal and vertical inter-channel time delays during a precedence-suppression task. A robust precedence effect was elicited via ear-level horizontal plane loudspeakers. In contrast, leading signal identification was minimal in the vertical condition and no systematic influence of the leading elevated and lowered median plane loudspeakers was witnessed suggesting that precedence was not active in the vertical condition. Observed influences that might have been generated by the lead-lag signal in the vertical plane was not consistent with any known precedence paradigms.

1 Introduction

Results in [1] found evidence that subject performance in the vertical domain may have been influenced by pinna-like elevation cues contained within, or created by, the lead-lag stimulus. Findings in [2] confirmed that assumption by separating the time- and spectrum-based influences during a precedence-based localization task in the horizontal, diagonal, and vertical parasagittal and median planes. This paper will report preliminary results from a third study focused on stimulus interactions across the horizontal azimuthal and median-sagittal planes.

2 Background

Auditory precedence is a well-researched two-loudspeaker product of a lead-lag signal paradigm. In this type of study, a leading signal is followed briefly with a replica signal from a second location. During a “precedence” event a listener is unaware of the

second signal and depending upon the amount of time between signals, the subject will localize the leading signal source and ignore the second signal. Within limits (~0.5 - 1.5 ms), greater time delay tends to proportionally increase the parasagittal (left or right of center) shift of the localized image [3] [4] [5].

The process works quite well in horizontal planes where binaural time and intensity cues (ITD and ILD) are strongest. However, in the median-sagittal plane the percept appears to be much weaker and results have been somewhat contradictory. Early studies found evidence of a median-sagittal “vertical” precedence [6] [7] while later studies investigating so-called “3-D” and “immersive” audio have noted a general lack of the effect [8] [9].

Utilizing stimuli and methods developed in [2], this study attempted to provoke directly comparable subject performance data across the transverse-parasagittal and median-sagittal planes. Outcomes

were expected to follow the experimental null hypothesis where lagging loudspeaker configuration does not affect lead-signal detection and therefore, localization was predicted to concentrate at the leading loudspeaker orientation and elevation.

3 Methods

Listening tests were executed in a 24 ft. x 24 ft. x 28 ft. anechoic chamber with additional anechoic wedges positioned to eliminate reflections from the floor. Five time-aligned loudspeakers were deployed with three placed across the ear-level (LVL) horizontal plane and two located in the elevated (ELV) and lowered (LWR) median-vertical plane. The loudspeakers were laser aligned and positioned along an arc 2.0 meters from the listening position and were time and level (62 dB-A SPL) calibrated to ensure proper incidence of the stimuli at the listening position. Four loudspeakers were located at 30° above, below, left and right of the center LVL loudspeaker. All loudspeakers were concealed behind acoustically transparent visually opaque curtains with a 5x5 response grid delineated on the curtain (see figure 1).

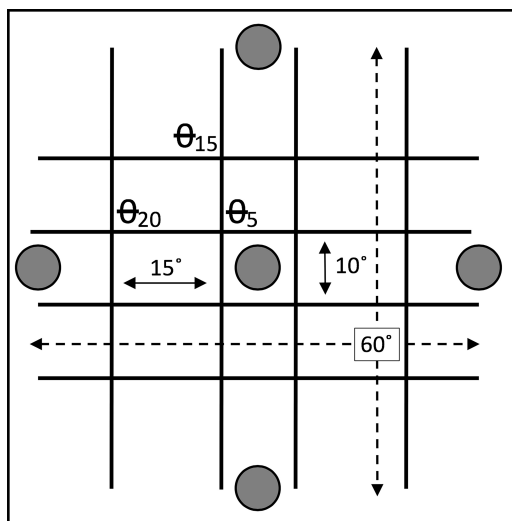


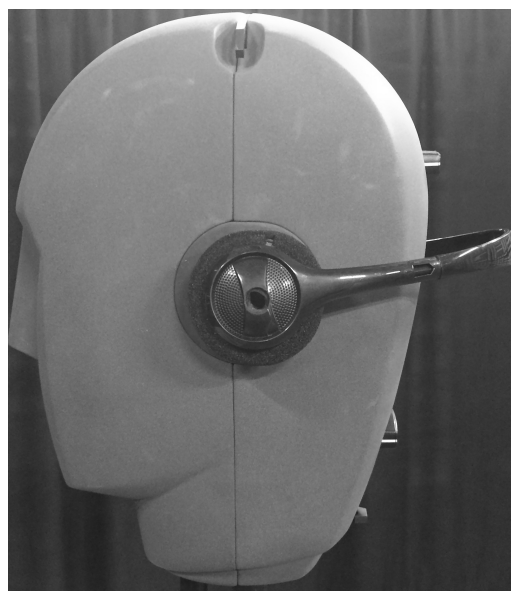
Figure 1. Depiction of the loudspeaker array and response grid.

Stimuli were a series of three, band-limited 500 ms noise bursts with 500 ms of silence between bursts. Two different noise bursts, one with high- and

lowpass (48-dB octave) filters set at 50 Hz and 15 kHz respectively and a second with the low-pass filter set at 5 kHz were employed. The former served as a reference with potential for robust pinnae cues and the latter functioned as an alternative where influential pinnae cues were not likely to be present.

Presentations were altered by changing the noise burst filter, the lead-lag configuration (horizontal or vertical) and the inter-channel time delay (ICTD) between loudspeakers, and by utilizing a head-worn device (“disruptor”) designed to acoustically alter the listener’s HRTF without changing the timing or overall spectral energy of the signals at the left and right ears of the listener.

The disruptor comprised a lightweight supra-aural stereophonic headphone set with loudspeaker drivers replaced with a small rubber tube about the diameter of the typical ear canal (photograph 2). A dense foam packing filled the cavity around the tube. Device design was based on the principal of direct concha excitation; the assumption that the posterior concha is a primary source of elevation-related reflections into the ear canal [9].



Photograph 1. Binaural head wearing the pinna-cue “disruptor” device.

The combination of two types of noise with the head-worn disruptor yielded three stimuli categories; normal (NORM) with unaltered pinnae cues, disrupted (DSRP) where individual pinnae cues were altered to become unfamiliar to the listener, and annihilated (ANIL) where energy above 5 kHz was removed. Figure 2 shows the FFT of impulse responses derived from sine-sweep recordings from the binaural head for ear-level NORM, DSRP, and ANIL stimuli. Note the markedly altered HRTF spectral characteristics of the signals above 2 kHz.

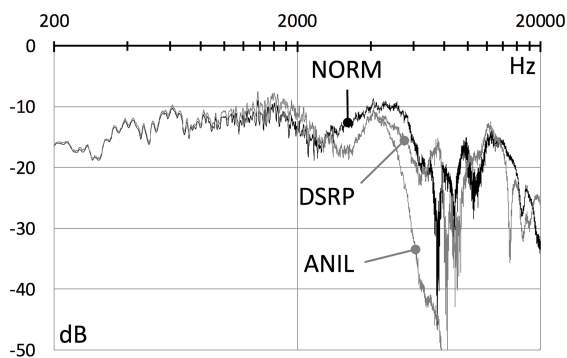


Figure 2. FFT of sine-sweep binaural recording of the ear-level stimuli.

The sequence of conditions was randomly assigned and stimuli were presented across three randomized sequences. Seventeen stimulus variations comprising five single loudspeaker presentations along with four horizontal and eight median-vertical lead-lag trials were presented with 0.5 and 1.0 ms ICTD. In total, each subject completed six trials for each stimulus type and generated 102 trials across 17 stimuli in each of the three conditions. Single loudspeaker parasagittal and median-sagittal trials were intermixed with horizontal and vertical lead-lag loudspeaker paired presentations. These stimuli served as controls for examination of the effect of the three stimuli types in isolation without the confluence of the delayed signal.

Subjects were not informed of the focus of the experiment and were simply asked to aim a laser pointer to the location from which they perceived the signal to originate. There were six seconds of silence

between each trial, allowing the listener sufficient time to indicate the perceived stimulus location. Localization percepts were noted and marked as having fallen within one of the zones delineated on the curtain. Zone hit count and percent correct identification of the lead loudspeaker orientation (far-left, left-of-center, center, right-of-center, far-right) and elevation (elevated, above-midline, ear-level, below-midline, lowered) were the dependent variables used to examine the influence of the stimuli in the parasagittal horizontal and median-sagittal planes.

Subjects were eight graduate audio engineering students with ages clustered in the mid-20s. All participants reported normal hearing and had completed a graduate course in critical listening. In a generalized sense, all subjects would be considered more experienced than “novice” but not yet “expert” listeners.

4 Results

Individual trials were counted as locus hits *per* the response grid shown in figure 1 and averaged across stimulus type categories to generate the dependent variables for analyses. *Shapiro-Wilk* tests indicated non-Gaussian distributions; thus, non-parametric Freedman ANOVA and *post hoc* Wilcoxon Matched Pairs tests were used to parse out the influence of the experimental stimuli and examine performance across group categories. Main effect *p*-values were not corrected for ties resulting in more conservative estimates of influence.

4.1 Control trials

Control stimuli comprised ear-level leftward, center, and rightward single-loudspeaker presentations (descriptive statistics are shown in table 1). Preliminary analyses indicated no influence of loudspeaker position ($\chi^2(2, n = 24) = 0.27, p = .873$). Thus, parasagittal and median data were pooled. ANOVA revealed a main effect of the NORM, DSRP, and ANIL conditions ($\chi^2(2, n = 24) = 18.5, p < .001$). *Post hoc* tests revealed NORM scores were significantly higher when contrasted with DSRP scores ($Z = 3.14, p = .002, r = .64, n = 24$) and ANIL ($Z = 4.04, p < .001, r = .82, n = 24$) scores. DSRP

scores were also significantly higher than ANIL scores, albeit, just barely ($Z = 2.02$, $p = .044$, $r = .41$, $n = 24$).

Table 1. Descriptive statistics for the LVL control presentations.

Statistic	NORM	DSRP	ANIL
<i>M</i>	.76	.53	.41
<i>Mdn</i>	.83	.58	.33
Range	.83	1.00	1.00
Skew <i>Z</i>	-2.66	-1.28	0.49
Kurt <i>Z</i>	4.29	2.87	2.15
<i>N</i> = 144	24	24	24
<i>Shapiro-Wilk</i> $W = 0.90$, $p < .001$, $N = 72$			

Consistent with prior work, results here show point-source perceptual elevation cues were systematically degraded by the disruptor and eliminated in the absence of spectral energy above 2 kHz.

4.2 Parasagittal ear-level presentations

Correct identification of the leftward of rightward orientation of the leading loudspeaker was greater than 97% across all three conditions. Preliminary analyses indicated no influence of leftward or rightward loudspeaker position ($Z = 1.46$, $p = .145$, $r = .30$, $n = 24$). Hence, parasagittal data were pooled by condition. ANOVA indicated a main effect of NORM, DSRP, and ANIL treatments ($\chi^2(2, n = 16) = 15.9$, $p < .001$). *Post hoc* tests revealed NORM scores were significantly higher than DSRP ($Z = 2.61$, $p = .009$, $r = .65$, $n = 16$) and ANIL scores ($Z = 3.42$, $p = .001$, $r = .85$, $n = 16$) while DSRP and ANIL scores were not significantly different ($Z = 1.45$, $p = .147$, $r = .36$, $n = 16$). Descriptive statistics are shown in table 2.

Table 2. Descriptive statistics for the parasagittal LVL (horizontal plane) presentations.

Statistic	NORM	DSRP	ANIL
<i>M</i>	.75	.49	.30
<i>Mdn</i>	.83	.58	.33
Range	.83	.83	.83
Skew <i>Z</i>	-2.00	-0.72	0.56
Kurt <i>Z</i>	3.80	1.88	2.02
<i>N</i> = 96	16	16	16
<i>Shapiro-Wilk</i> $W = 0.90$, $p < .001$, $N = 48$			

Results here confirmed the ICTD treatment elicited a robust measurable precedence effect across the horizontal plane. Here again, perceptual elevation cues were systematically degraded by the disruptor and eliminated in the absence of spectral energy above 2 kHz reaffirming the efficacy of the disruptor for removing the influence of the pinna.

4.3 Median-sagittal presentations

Examination of the vertical authority of the ICTD revealed no reliable directional influence of the time-based stimuli. ICTD effect was tabulated as the hit count for the predicted perceptual lead-lag locus direction. ELV 1.0 ms ICTD stimuli generated 49% mean hit count for the DSRP and ANIL conditions while LWR stimuli generated 35% mean hit count. In like manner, ELV 0.5 ms ICTD stimuli generated 15% mean hit count for the DSRP and ANIL conditions while LWR stimuli generated 20% mean hit count.

Detailed in tables 3 and 4 are the perceptual response counts for targets as mapped to the vertical response grid depicted in figure 1. Here, darker cells represent greater hit numbers ($n = 96$ trials per stimulus type).

Table 3. Percept response count by condition, ICTD, and elevation for median-sagittal ELV target stimuli.

COND	NORM		DSRP		ANIL	
ICTD ms	1.0	0.5	1.0	0.5	1.0	0.5
Percept Response Counts						
+15-30°	33	7	17	9	23	7
+5-15°	25	18	25	4	30	9
LVL	18	25	34	37	31	40
-5-15°	7	11	14	26	10	27
-15-30°	13	35	6	20	2	13

Table 4. Response count by condition, ICTD, and elevation for median-sagittal LWR-target stimuli.

COND	NORM		DSRP		ANIL	
ICTD ms	1.0	0.5	1.0	0.5	1.0	0.5
Percept Response Counts						
+15-30°	15	5	7	16	15	21
+5-15°	21	20	14	24	15	18
LVL	20	37	44	28	28	27
-5-15°	11	14	21	14	29	23
-15-30°	29	20	10	13	9	7

Statistical analyses indicated no measurable differences between the ELV and LWR scores for the 0.5 and 1.0 ms stimuli ($Z_{1.0\text{ms}} = 1.17, p = .242, r = .17, n = 48$; $Z_{0.5\text{ms}} = 1.23, p = .217, r = .17, n = 48$). ANOVA also revealed no main effect of the NORM, DSRP, and ANIL conditions for both ICTD categories (1.0 ms ELV: $\chi^2(2, n = 16) = 0.65, p = .720$; 0.5 ms ELV: $\chi^2(2, n = 16) = 2.09, p = .351$; 1.0 ms LWR: $\chi^2(2, n = 16) = 1.22, p = .544$; 0.5 ms LWR: $\chi^2(2, n = 16) = 3.41, p = .182$).

5 Discussion

The precedence effect has been characterized as “binaurally-mediated” across transverse planes and “spectrally-mediated” in the median-sagittal plane [6]. Previous works have established precedence-based outcomes to be robust across the horizontal-azimuthal frontal planes where clear binaural cues are present. In the median-sagittal plane where binaural cues are minimized, reliable observations have been somewhat elusive and contradictory.

In the ear-level parasagittal trials, DSRP and ANIL conditions both generated debilitating cues for the detection of elevation such that percept scores were not found to be statistically different between these two conditions. The level of observed degradation was to an extent that subjects were essentially guessing when the loudspeaker position was in what one would anticipate to be the best possible location, i.e., ear-level horizontal plane. Given, as shown in figure 2, overall spectral energy was comparable between NORM and DSRP stimuli, results strongly suggest it is not just the presence of upper-band spectral energy, but rather, spectral energy plus individualized HRTF-generated cues that govern the detection of a sound source’s elevation during a precedence event.

In the present study, robust binaural mediation for identification of the parasagittal direction of the leading signal was observed in the horizontal plane. However, without reliable and familiar spectral-based cues, subjects were unable to identify the exact location of the source. If this is the case, one must extrapolate that observation to precedence in the median-sagittal plane and accept a similar process to

be active during a vertically oriented precedence-based event.

In the median-sagittal presentations, the NORM 1.0 ms ICTD elevated and lowered conditions appear as one might expect in a precedence experiment where a majority of the percepts accumulated in the direction of the leading loudspeakers (tables 3 and 4). Results for the NORM stimuli here could give an impression, albeit a weak one, of precedence-like activity in the vertical plane. However, percepts across the other stimuli reveal a curious non-systematic distribution of elevation responses; where, disrupted (DSRP) and annihilated (ANIL) elevated- and lowered-leading presentations show percepts accumulated mostly around the ear-level regions in both ICTD categories. Results here failed to elicit any systematic directional cues *via* the ICTD similar to those observed with horizontal plane stimuli.

To examine the interaction of the lead-lag paradigm with spectral influences from the HRTF, perceptual data are displayed in figures 3 and 4 for the three stimuli conditions (see next page). Here, top and bottom panels plot respectively percepts for the elevated- and lowered-leading loudspeaker conditions. Boxes indicate leading loudspeaker target locations. Each dot represents one percent of the responses for its respective stimuli and condition. Dots in and around a boxed area signify target identifications and those not in boxed areas represent locations of elevation percepts across the vertical response grid as depicted in figure 1. Percent magnitude is cited for each elevation.

While the 1.0 ms ICTD NORM condition shows a concentration of percepts at the leading loudspeakers suggesting a possible influence of precedence (figure 3), DSRP and ANIL stimuli across figures 3 and 4 confirm a randomized distribution of percepts spanning the median plane. For the 0.5 ms ICTD stimuli, even when the upper loudspeaker was leading, percepts across all three conditions appeared to accumulate in the lowered regions (figure 4, top panel), suggesting the lagging stimuli suppressed the leading signal, an observation not consistent with known precedence outcomes. For the lower-leading

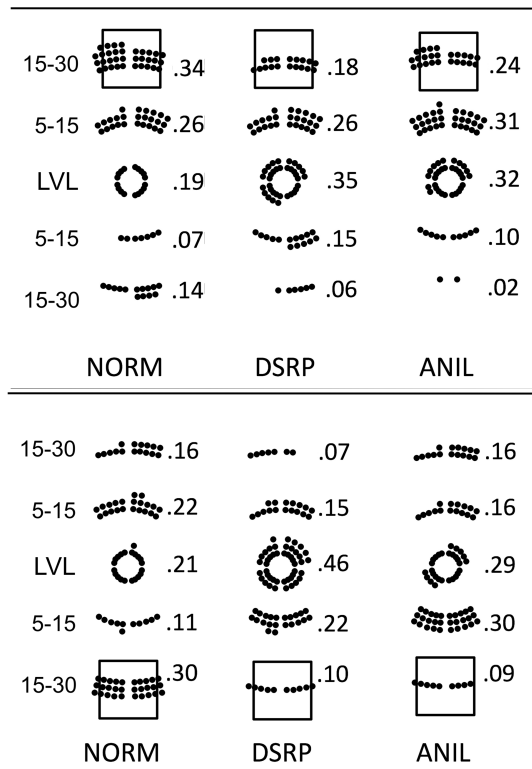


Figure 3. Vertical percepts for the 1.0 ms ICTD.

0.5 ms ICTD stimuli (figure 4, lower panel), percept responses are generally distributed across the median plane in all three conditions. In both ICTD categories, there are no recognizable patterns of note that might suggest adherence to a lead-lag signal paradigm when the spectral information was altered or removed. In the 1.0 ms NORM condition (figure 3, leftward column), it is a reasonable possibility that the lead time was just long enough to provide enough salient spectral information for the system to gain some minimal, albeit not very accurate, directional elevation cues. However, any elevation cues gained in that brief moment were immediately wiped out when the spectral information was altered or eliminated in the DSRP and ANIL conditions respectively.

In the median-sagittal plane, had either time or spectrum significantly dominated the vertical percept, the former would have resulted in target hits even after spectral cues were degraded. On the other hand,

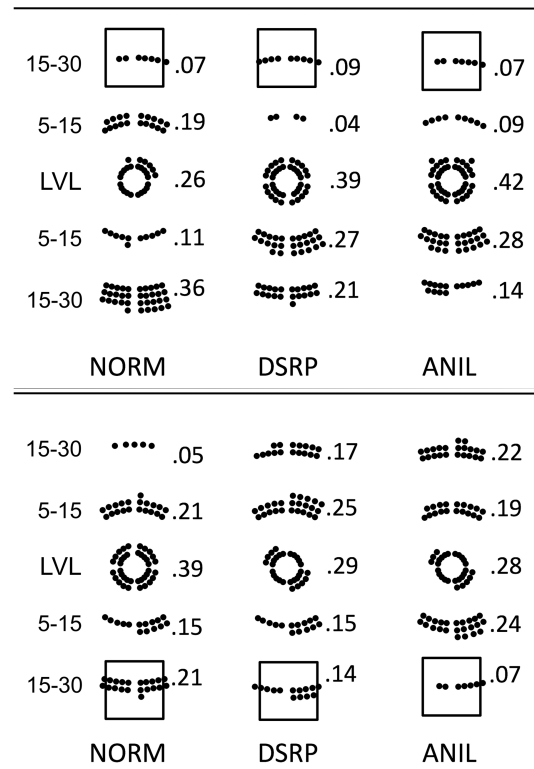


Figure 4. Vertical percepts for the 0.5 ms ICTD.

if precedence in the median-sagittal plane is simply spectrally-mediated as described in prior works, the latter should have produced positive responses across the altered pinnae-cue condition. Instead, altered spectral cues were essentially the same as no spectral cues, suggesting again, it was spectral energy *plus* individualized HRTF-generated cues that governed the detection of stimuli elevation. Results here provide clear and powerful evidence that after alteration of the spectral energy associated with HRTF generated cues, the remaining time-based cues had no influence on the vertical percept of the lead-lag stimuli.

This experiment altered and eliminated the spectral influences that govern the detection of stimulus elevation and presented two different horizontal and vertical ICTDs during a precedence-suppression task. A robust precedence effect was elicited *via* ear-level horizontal plane loudspeakers. In stark contrast, leading signal identification was minimal in the

vertical condition and no systematic influence of the leading elevated and lowered median-plane loudspeakers was observed.

6 Conclusions

This experiment attempted to elicit comparable auditory precedence in the horizontal and vertical planes. Methods employed here were based on proven models where robust precedence-based suppression effects have been provoked and observed.

Performance measures for the horizontal plane condition indicated robust precedence was observed. In contrast, in the median-plane condition, no recognizable vertical locus percept patterns emerged.

Results here advocate that any identification of the leading loudspeakers in the median-sagittal plane was likely the sole product of individually unique spectrally-based HRTF generated influences.

So-called “precedence” comparable to that observed in the horizontal-binaural domain was not found in the vertical condition.

References

- [1] P. Mayo and W. Bulla, "Lead-signal localization accuracy for inter-channel time difference in higher and lower vertical, side, and diagonal loudspeaker configurations," in *143rd AES Convention* (<http://www.aes.org/e-lib/browse.cfm?elib=19229>), New York, NY, USA, 2017.
- [2] W. Bulla and P. Mayo, "The effect of pinnae cues on lead-signal localization in elevated, lowered, and diagonal loudspeaker configurations," in *AES 145th Convention, paper #10066*, <http://www.aes.org/e-lib/browse.cfm?elib=19792>, New York, NY, USA, 2018.
- [3] A. D. Brown, G. C. Stecker and D. J. Tollin, "The precedence effect in sound localization," *JARO*, vol. 16, pp. 1-28 (DOI: 10.1007/s10162-014-0496-2), 2015.
- [4] N. Viemeister, "An overview of psychoacoustics and auditory perception," in *AES 8th International Conference, The Sound of Audio*, 1990.
- [5] H. Wallach, E. Neumann and M. Rosenzweig, "The precedence effect in sound localization. Reprint from Am. J. Psych, 52, 315-336 (1949)," *J. Audio Eng. Soc.*, vol. 21, no. 10, pp. 817-826, 1973.
- [6] R. Litovsky, B. Rakerd, T. Yin and W. Hartmann, "Psychophysical and physiological evidence for a precedence effect in the median sagittal plane," *J. Neurophysiology*, vol. 77, no. 4, pp. 2223-2226, 1997.
- [7] M. Dizon and R. Litovsky, "Localization dominance in the median-sagittal plane: effect of stimulus duration," *J. Acoust. Soc. Am.*, vol. 115, no. 6, pp. 3142-3155, 2004.
- [8] L. Davis and H. Lee, "Echo thresholds for a 3D loudspeaker configuration," in *AES 140th Convention*, Paris, France, 2016.
- [9] R. Wallis and H. Lee, "Vertical Stereophonic Localization in the Presence of Interchannel Crosstalk: The Analysis of Frequency-Dependent Localization Thresholds," *J. Audio Eng. Soc.*, vol. 64, no. 10, pp. 762-770 (DOI: <https://doi.org/10.17743/jaes.2016.0039>), October 2016.
- [10] H. Lee and F. Rumsey, "Level and time panning of phantom images for musical sources," *J. Audio Eng. Soc.*, vol. 61, pp. 978-988, 2013.